

# VIDEO AND DIGITAL MULTIMEDIA AGGREGATOR

## CONTENT SUGGESTION ENGINE

### Related Applications

This application is a continuation-in-part of U.S. patent application serial number 08/906,469, filed August 5, 1997, entitled "Reprogrammable Terminal for Suggesting Programs Offered on a Television Program Delivery System," which is a continuation of U.S. patent application serial number 08/160,281 filed December 2, 1993, entitled "Terminal for Suggesting Programs Offered on a Television Program Delivery System," now U.S. Patent 5,798,785, which is a continuation-in-part of U.S. patent application serial number 07/991,074, filed December 9, 1992, entitled "Remote Control for Menu Driven Subscriber Access to Television Programming." The above patent applications and patent are hereby incorporated by reference.

The following U.S. Patents also are incorporated by reference:

U.S. Patent 5,798,785, entitled "Terminal for Suggesting Programs Offered on a Television Program Delivery System;" U.S. Patent 5,986,690, entitled "Electronic Book Selection and Delivery System;"

The following co-pending patent applications also are incorporated by reference: patent application serial number 09/191,520, filed November 13, 1998, entitled "Digital Broadcast Program Ordering;"

patent application serial number 09/289,957, filed April 13, 1999, entitled "Electronic Book Alternative Delivery Systems;"

patent application serial number 09/289,956, filed April 13, 1999, entitled "Electronic Book Alternative Delivery Methods;" and

patent application entitled "Video and Digital Multimedia Aggregator;" and patent application entitled "Video and Digital Multimedia Aggregator Remote Content Crawler," filed on date herein.

### Technical Field

The technical field is television program and digital multimedia delivery systems that incorporate intelligent and flexible program search and delivery mechanisms.

## Background

Cable television delivery systems greatly expanded the array of programs available to television viewers over that provided by over-the-air broadcast systems. Subsequently, satellite television delivery systems, and in particular, digital satellite broadcast systems further expanded the viewing choices for consumers. In the near future, digital broadcast television systems will provide many more programming choices for consumers.

In addition to television programming delivered through television program delivery systems, other programs and events may be available to consumers. These other programs and events include streaming video sent over wired and unwired, narrowband to broadband services, digital audio programs, and other multimedia data.

Unfortunately, customers are still limited in their television viewing choices by the local and regional nature of television delivery systems. For example, a broadcaster in Boston may provide programming of local interests to people in Massachusetts while a broadcaster in Seattle may provide different programming to people in the Seattle area. A person in Boston generally cannot access Seattle programming, other than programming that is provided at a national level.

In addition to this local/regional market segregation, many other sources of programming and events may not be available in a specific viewing area. These other sources may include audio programming, streaming video, local or closed circuit television programming (e.g., education television programming provided by a state education department) and other programming.

Even if certain programming is available in a local viewing area, a viewer may not be aware of its existence. This situation may be the result of a large array of available programming coupled with a limited program menu or guide. The program guide may be limited in that not all available programming can be listed, some programming changes occur that are not reflected in the program guide, and errors may exist in the program guide. In addition, the more comprehensive the program guide, the harder it is for the viewer to search and navigate the program guide to find a desired program.

## Summary

The problems noted above are solved by the video and digital multimedia aggregator system and method described herein. Program content can be packaged and delivered by the

1 system, including video, television, radio, audio, multimedia, computer software, and  
2 electronic books, or any content that can be delivered in digital format.

3 The aggregator comprises a request and results processing server, a search engine  
4 server coupled to the request and results processing server and a content acquisition server  
5 coupled to the request and results processing server. The request and results processing  
6 server receives a request for a program, the search engine server searches for the program  
7 and the content acquisition server acquires a program for delivery to the user. The request  
8 and results processing server includes a search request processor that receives information  
9 related to the user's search request and provides the information to a search results form  
10 builder that creates an electronic search request. The search request may be augmented by  
11 using a content search suggestion engine to add additional search terms and descriptions to  
12 the search request. The aggregator may also include a decoder that decodes program content  
13 and program metadata from remote sources for storage at the aggregator, and an encoder that  
14 encodes content metadata and programs for delivery to the user. The aggregator may also  
15 comprise one or more crawlers, such as a content crawler, to look for program content in the  
16 digital communications network.

17 The search engine server searches at least a local content database. The local content  
18 database comprises at least two file types. A content file includes a complete program  
19 content file. For example, the 1997 movie *Titanic* may exist in the local content database as  
20 a complete program content file. The complete program content file may also include a  
21 reference file or metadata that contains additional information related to the content.  
22 Alternatively, the metadata may be contained within the complete program content file itself.  
23 Such additional information in the reference file may include: a program description,  
24 including program rating, program description, video clips, program length, format (e.g., 4x3  
25 television or 16x9 movies), and other information; billing information and digital rights  
26 management information; viewing statistics, including number of times viewed, dates/times  
27 viewed, identity of users viewing the program; advertisement information to allow ads to be  
28 inserted during viewing of the program; and other information.

29 The additional information in the reference file and/or the complete program content  
30 file may be provided in whole or in part to the system users. For example, the aggregator  
31 may provide a program description and accompanying video clips to selected systems users.

1 The reference file and/or the complete program content file may also be used by the  
2 aggregator for system administration purposes. For example, billing and digital rights  
3 management information may be used to collect appropriate fees from system users and to  
4 provide such collected fees to the entities owning rights in the content.

5 A remote content crawler continually crawls the digital communication network  
6 looking for content to provide to the aggregator. The content provided to the aggregator may  
7 be stored in a form of an entire content file. For example, the content may include an entire  
8 movie, television program or electronic book. Alternatively, the content provided to the  
9 aggregator may be a reference to a content file that is stored at, or that will be available at  
10 one of the remote locations. For example, the content may be a reference to a future,  
11 scheduled live sports event that will be made available to system users. The sports event  
12 may be provided for a one time fee, as part of a sports package, for which a fee is collected,  
13 or as a free event. In the examples discussed above, the content may be stored at the  
14 aggregator, and may subsequently be provided to system users. For the example of the live  
15 sports event, the aggregator may store the live sports event and may then provide the sports  
16 event as a replay, in addition to facilitating live viewing of the sports event.

17 To allow a user to maximize use of the video and digital multimedia aggregator, a  
18 search suggestion engine may be used to provide suggestions of available content. The  
19 search suggestion engine starts by constructing a database of metadata elements or word  
20 items. The metadata element database may be constructed by using a crawler that  
21 periodically or continually crawls a content metadata database and extracts word items from  
22 the content metadata database. The word items are then processed according to a number of  
23 factors, including metadata field description, part-of-speech, entries in dictionaries and  
24 thesaurus and other factors. The thus-processed word items are indexed, and may be  
25 provided with a vector value. A subsequent search request submitted by the user may  
26 include search criteria having vector values. A processor within the search suggestion engine  
27 compares the vector values to determine if an indexed word item should be included as a  
28 suggested search term to be appended to the user's search criteria. A list of suggested  
29 content is then generated. The list of suggested items may be ranked and filtered before  
30 being provided to the user.

Specifically, in an embodiment, an apparatus for suggesting available content in a digital communications network, comprises a content metadata crawler that searches metadata related to the available content and produces a metadata list, wherein the metadata list comprises a plurality of metadata elements, and wherein each metadata element comprises one or more metadata fields; a suggestion keyword indexer coupled to the content metadata crawler, wherein the suggestion keyword indexer receives the metadata list and indexes the metadata elements; a suggestion database coupled to the suggestion keyword indexer that stores the indexed metadata elements; and a suggestion database processor coupled to the content metadata crawler, the suggestion keyword indexer and the suggestion keyword database, wherein the suggestion database processor searches the suggestion database, based on one or more search request criteria, to produce a list of suggested search criteria that may be used to identify suggested content. The suggestion keyword indexer, comprises an extraction module that extracts and caches a value of each metadata field; a parsing module coupled to the extraction module that parses contents of uniquely identifying metadata fields, wherein the contents of a uniquely identifying field comprises one or more word items; a classifying module coupled to the parsing module that classifies one or more of the one or more word items; a comparison module coupled to the classifying module that compares one or more of the one or more word items to determine a list of related terms; and an index matrix record builder that creates and augments an index matrix record for each of the classified word items.

The apparatus may also include a dictionary database, a thesaurus database and a lexicon database. The comparison module compares a word item to entries in the dictionary database, the thesaurus database and the lexicon database, and the list of related terms includes one or more of a dictionary definition, lexicon data, and one or more synonyms.

The classifying module may comprise one or more computational linguistics tools, such as a rule-based part-of-speech tagging algorithm and a stochastic part-of-speech tagging algorithm, for example. The computational linguistic tools may be used to determine part-of-speech data of a word item. The index matrix record builder then adds the part-of-speech data to the index matrix record for the word item.

1 The uniquely identifying fields comprise one or more of content type, content title,  
2 date of production, rating and parental notice information, performer, artist, writer, author,  
3 plot summary, keyword list, and textual content description.

4 The index matrix builder may comprise a vector assignment module that assigns a  
5 word item vector value for a word item, wherein the word item vector value is a measure of  
6 similarity between a word item and a related term.

7 The suggestion database processor may comprise a vector determination module that  
8 assigns a search term suggestion vector range to one or more of the search request criteria;  
9 and a vector value comparator that compares the search term suggestion vector range and the  
10 word item vector value to determine if the word item vector value falls within the suggestion  
11 vector range, wherein word items that fall within the suggestion vector range are used to  
12 search the suggestion database. The suggestion vector range may be adjustable by a user of  
13 the apparatus. For example, the user may change a setting for relative breadth of the  
14 suggested content where a narrow setting may result in fewer, more-relevant suggested  
15 content items and a broader setting may result in a larger number of broader results.

16 The apparatus may further comprise a user-defined filter, including a user history  
17 filter; a user profile filter; and an approved content access filter, wherein the suggestion  
18 database processor processes search results from the suggestion database using the user-  
19 defined filter to produce the list of suggested content. The apparatus may also include a  
20 ranking module, wherein the ranking module ranks content in the list of suggested content.  
21 The ranking module may rank the content according to one or more of a user historical  
22 analysis report and similarities to previously accessed content by the user.

23 A method corresponding to the above apparatus may include the steps of constructing  
24 a database of indexed metadata elements; receiving a content search request from a user of a  
25 digital communications network; comparing the search request to the database of indexed  
26 metadata elements; caching indexed metadata elements that satisfy the search request;  
27 retrieving a user profile for the user; filtering the cached metadata elements according to the  
28 user profile; ranking the filtered metadata elements; and providing the ranked metadata  
29 elements to a search request processor as criteria for returning suggested content.

30 The step of constructing the database of indexed metadata elements may include  
31 opening one or more metadata records in the content metadata database; for a current one of

1 the one or more metadata records, determining if end-of-file has been reached, reading an  
2 entire metadata entry of the current metadata record, wherein the current metadata record  
3 comprises one or more uniquely identifying fields, and wherein each of the one or more  
4 uniquely identifying fields comprises one or more terms, extracting and caching a value for  
5 each term for one or more of the one or more uniquely identifying fields, and parsing and  
6 caching terms of each of the uniquely identifying fields. The method may further comprise,  
7 for each cached term, determining if an index record exists for the cached term; and if no  
8 index record exists, creating an index matrix record, and adding the cached value to the index  
9 matrix record. The step of creating the index matrix record may comprise determining a part  
10 of speech of the term in order to identify additional part of speech data; comparing the term  
11 to thesaurus data to determine similar terms, and storing the part-of-speech data and the  
12 similar terms as the index matrix record.

### 13 **Description of the Drawings**

14 The detailed description will refer to the following drawings in which like numerals  
15 refer to like elements, and in which:

16 Figure 1 is a block diagram of primary components of a content search, packaging,  
17 and delivery system;

18 Figure 2 is a block diagram of the components of the content search, packaging, and  
19 delivery system configured to deliver content to a set top terminal;

20 Figure 3 is a block diagram of the components of the content search, packaging, and  
21 delivery system configured to deliver content to a personal computer terminal;

22 Figure 4 is a schematic of the components of the content search, packaging, and  
23 delivery system showing subsystems of an aggregator, a user terminal, and a remote content  
24 server;

25 Figure 5 is a schematic of subsystems of a request and results processing server and  
26 components with which the processing server interacts as part of the content search,  
27 packaging, and delivery system;

28 Figure 6a is a schematic of subsystems of a content search suggestion engine and  
29 components with which the content search suggestion engine interacts as part of the content  
30 search, packaging, and delivery system;

Figure 6b is a diagram of components of a suggestion keyword indexer used with the search suggestion engine of Figure 6a;

Figure 6c is a diagram of components of a suggestion database processor used with the search suggestion engine of Figure 6a;

Figure 7 is a schematic of subsystems of a database administrator as part of the content search, packaging, and delivery system;

Figure 8 is a schematic of subsystems of a system administrator and components with which the system administrator interacts as part of the content search, packaging, and delivery system;

Figure 9 shows a flowchart describing a process of building and maintaining a database of suggestion terms;

Figure 10 shows a flowchart describing a process of using current search request criteria along with user profile and history data for retrieving suggested search parameters from the suggestion database;

Figures 11a–11c show a search request form; and

Figures 12a-12b show a search results form.

## **DETAILED DESCRIPTION**

Figure 1 is a block diagram of a content search, packaging, and delivery system 200. The content to be packaged and delivered by the system 200 includes video, television, radio, audio, multimedia, computer software and electronic books. Components of the system 200 include an aggregator 201 and a user terminal 202, which are connected using a wide area distribution system 203. Other components are remote content servers 204 that exchange data with the aggregator 201 using a wide area network/Internet 205 connection. The user terminal 202 may incorporate a video display system 207, an audio playout system 208, an audio/video recording system 209, user input devices 214, an electronic book reader 230, and a connection to a home network 229 to interact with other devices in the user's home environment. Alternatively, one or more or all of the video display system 207, the audio playout system 208, the audio/video recording system 209, and the electronic book reader 230 may be separate components that are coupled to the user terminal 202.

The system 200 allows a user to enter search parameters, such as keywords and category headings, and have the aggregator 201 use these parameters to locate, package, and



1 deliver content to the user terminal 202 from numerous sources. The requests and content  
2 deliveries can be sent over communications links including, but not limited to, telephone  
3 lines, coaxial cable, fiber-optic cable, wireless connections, wide area networks, the Internet,  
4 and other communication media collectively represented by the wide area distribution system  
5 203. The numerous sources of content are shown in Figure 4 and include, but are not limited  
6 to, an aggregator local storage 254, local streaming sources 262, remote content storage 258,  
7 and remote streaming sources 259. In an embodiment, the local streaming sources 262 are  
8 comprised of the audio and video channels being delivered using a cable television headend  
9 system that may also house the aggregator 201.

10 The system 200 will take a user's search request and may perform a search of  
11 virtually every active and scheduled radio and television source in the world, as well as  
12 archived sources of video and audio programming such as movies and recorded audio  
13 sources and sources of other multimedia, software, and electronic book content. In an  
14 embodiment, the system 200 will also search Internet Web sites and other online databases.  
15 The user will then be able to select programming or other content for download based on the  
16 search results. In an embodiment, the download, or delivery, process can be fulfilled by  
17 making the content available on a specific channel of a cable television system, or by  
18 transmitting the content using a digital communications protocol, such as the Internet  
19 standard TCP/IP, for example. In addition, the system 200 is capable of formulating and  
20 presenting a list of suggested content based on an analysis of the user's current search  
21 parameters, stored information about previous searches and previously selected content  
22 downloads and other user-specific or related information. The system 200 is also capable of  
23 notifying a user prior to the start time of selected programming and availability of other  
24 content using such notification as an electronic mail message and/or an on-screen message  
25 indicating that the scheduled program will be broadcast at a specified time. The system 200  
26 may support one or more digital rights management (DRM) systems to track the usage and  
27 copyrights associated with downloaded content and bill the user's account as appropriate and  
28 provide any license and usage fees to the content provider. The system 200 may also  
29 implement a users' privacy protection scheme allowing users to control what information is  
30 gathered, limit what is done with that information, and review and delete a user's profile if  
31 desired.

1 An overview of an embodiment of the system 200 is shown in Figure 2 where the  
2 user terminal 202 is a set top terminal 206 that communicates with the aggregator 201  
3 through a cable television system headend 210, thereby making use of the cable television  
4 system headend 210 high bandwidth concatenated cable system 216. The set top terminal  
5 206 system may include a tuner 228, a demultiplexer 231, the video display system 207, the  
6 audio playout system 208, the audio/video recording system 209, user input devices 214, an  
7 electronic book reader 230, and a connection to the user's home network 229. These  
8 components are used to tune, select, view, listen to, and store audio and video programming  
9 and other content delivered by the system 200. Figure 2 also shows a communications path  
10 from one or more remote content servers 204 through the wide area network/Internet 205  
11 directly to the set top terminal 206, which bypasses the aggregator 201 and cable television  
12 system headend 210. This path may be used in the case where the requested content is  
13 available in the required format from the remote content server 204 and is authorized for  
14 direct delivery to the user. In an alternative embodiment, the aggregator 201 is collocated  
15 with the cable television system headend 210 that is acting as the wide area distribution  
16 system 203 as is the case when a cable television system is also serving as the user's Internet  
17 service provider.

18 An overview of another embodiment of the system 200 is shown in Figure 3, where  
19 the user terminal 202 is a personal computer terminal 211 that communicates with the  
20 aggregator 201 through an Internet service provider/cable television system headend 215. In  
21 this case, the content may be delivered by a cable television headend that is operating as an  
22 Internet service provider (ISP). The personal computer terminal 211 may also include the  
23 video display system 207, the audio playout system 208, the audio/video recording system  
24 209, user input device 214, the electronic book reader 230, and the connection to the home  
25 network 229. A communications path may also exist from one or more remote content  
26 servers 204 through the wide area network/Internet 205 directly to the personal computer  
27 terminal 211. This communications path bypasses the aggregator 201 and the Internet  
28 service provider/cable television system headend 215 in the case where requested content is  
29 in the required format and is authorized for direct delivery to the user.

30 The user can receive video and audio programs (i.e., the content) in a number of  
31 different ways. First, the tuner 228 in the user terminal 202, shown in Figure 2, can simply

1 tune to the desired program, or the demultiplexer 231 can select a program from a digital  
2 multiplex thereby displaying the program at the appropriate time. However, the desired  
3 program may not be broadcast by the user's cable television system headend 210, or the user  
4 may want to watch a program that is supplied by the local headend, but not subscribed to by  
5 the user, e.g., a boxing match broadcast by a premium channel that the user does not  
6 ordinarily receive. In these examples, the program can be delivered to the user by the  
7 aggregator 201 using telephone lines, fiber-optic transmission lines, or other communication  
8 media, or using the cable television system headend 210. In an embodiment, the aggregator  
9 201 can supply programs by pulling programs from program delivery systems in the United  
10 States and abroad and delivering the programs to system users. The aggregator 201 can,  
11 therefore, provide a user access to programs outside the user's normal viewing area.

12 Figure 4 shows the system 200 and system components in more detail. At a user's  
13 location, the user terminal 202 includes the tuner 228, the demultiplexer 231, a user terminal  
14 processor 227, user local storage 212 and user local database 213. Coupled to the user  
15 terminal 202 may be the user input devices 214, the video display system 207, the audio  
16 playout system 208, the audio/video recording system 209, an electronic book reader 230,  
17 and a connection to the home network 229.

18 The user terminal 202 may be coupled through the wide area distribution system 203  
19 to the aggregator 201 and further through the wide area network/Internet 205 to remote  
20 program sources. The remote program sources include the remote streaming sources 259 and  
21 the remote central storage 258. The remote program sources also include remote databases  
22 261 and, through the remote content server 204, a remote server database 260.

23 The aggregator 201 may include a communications server 250 that communicates  
24 with the user terminal 202 through the wide area distribution system 203. The  
25 communications server 250 receives inputs from a request and results processing server 300,  
26 a content delivery server 450 and a system administrator 500. The content delivery server  
27 450 receives inputs from a coder and content formatter 253 and a content acquisition server  
28 400. The content delivery server 450 also accesses the aggregator local storage 254 and the  
29 local streaming sources 262. Finally, the content delivery server 450 provides an output to  
30 the system administrator 500.

1 The coder and content formatter 253 receives inputs from the content acquisition  
2 server 400, the aggregator local storage 254 and the local streaming sources 262. The system  
3 administrator 500 receives inputs from the content and delivery server 450, and  
4 communicates with the content acquisition server 400, the request and results processing  
5 server 300, a search engine server 350, and aggregator archives 255. A decoder and content  
6 formatter 252 is coupled to the content acquisition server 400. Finally, a network gateway  
7 251 couples components of the aggregator 201 with the remote content server 204 through  
8 the wide area network/Internet 205.

9 Programs received at the aggregator 201 may be input to the formatter 253. The  
10 formatter 253 reformats, as necessary, all input content into a format that is readily received  
11 by the user terminals 202 operably connected to the system 200. In addition, the formatter  
12 253 can store full or partial copies of content in compressed form in the aggregator local  
13 storage 254. The aggregator 201 can provide real-time delivery of certain content, such as a  
14 boxing match. In an embodiment, if a large number of users want a particular live program,  
15 then the cable television system headend 210 (see Figure 2), or similar program source, can  
16 broadcast the program on a particular channel available to all the requestors instead of  
17 broadcasting the program to each individual user over the wide area distribution system 203.  
18 In addition to delivery from a cable television headend, content may be provided by a  
19 satellite delivery system, over-the-air-broadcast, and through the Internet. See U.S. Patent  
20 Application serial number 09/191,520, entitled Digital Broadcast Program Ordering, hereby  
21 incorporated by reference, for additional details of broadcast program delivery.

22 The aggregator 201 can also implement a screening process for limiting the number  
23 of programs captured to those programs with a viewing audience above a predetermined  
24 threshold. The aggregator 201 may contain a filter that will pass only programs meeting the  
25 predetermined selection criteria. The filter may include programming that screens the  
26 content to reject specific items, such as adult content, for example.

27 The system administrator 500 records what fees should be paid and to whom. For  
28 example, the aggregator 201 will determine to whom any copyright or other fees should be  
29 paid when a program is broadcast.

30 The user terminal 202 may be a television, a set top terminal 206, a personal  
31 computer terminal (not shown), or any wired or wireless device capable of receiving digital

1 or analog data, or some combination thereof. The user terminal 202 is equipped with the user  
2 input devices 214 that communicate search criteria to the system 200 as well as navigate  
3 through the user terminal menu system and control the user terminal's other various  
4 functions. The user local storage 212 is used to store and archive content onto one or more  
5 removable and/or non-removable storage devices or media for later access. Removable  
6 storage media may include, but is not limited to, magnetic tape, magnetic disks, optical disks  
7 and modules, and electronic memory cartridges. The user local database 213 may store  
8 relevant information about a user's profile and account. This information includes, but is not  
9 limited to, user name, password, personal information that the user has authorized for  
10 storage, billing information, other users allowed access to the account, past search criteria,  
11 past content download information, and library information about stored content. As a  
12 consumer protection device, the user terminal 202 may enable the user to view the  
13 information stored in the user local database 213 and modify certain data fields and select  
14 which data fields may be reported to a main system database (not shown) within the  
15 aggregator 201. Certain fields including, but not limited to, account numbers and billing  
16 information may not be allowed this level of user access.

17 The user terminal processor 227 may include a central processing unit and all  
18 associated hardware, software, and firmware modules to perform all operations within the  
19 user terminal 202. These operations include, but are not limited to, managing  
20 communications with the aggregator 201 and other networked devices, processing search and  
21 download requests, displaying search and download results, managing and controlling  
22 communications with the user local storage 212 and the user local database 213, responding  
23 to user interaction with presentation of graphical user interface (GUI) menus, playing out  
24 selected programming content using various audio and video output devices, implementing  
25 the user's part of the digital rights management schema, and administering the user's account  
26 and billing. The tuner 228 and the demultiplexer 231 are used to select an audio/video  
27 channel for playout from the channels available on the cable television system 216.

28 In an embodiment, the user terminal 202 may incorporate selected features of the  
29 aggregator 201. For example, the user terminal 202 may include a metadata crawler, an  
30 aggregator, and program content and program metadata storage.

1 The user terminal 202 communicates with the aggregator 201 using the wide area  
2 distribution system 203. Within the aggregator 201, the communications server 250 acts as  
3 the interface point to the wide area distribution system 203 for the purpose of managing  
4 primary communications to system users. The communications server 250 routes incoming  
5 user requests and associated user information to the request and results processing server  
6 300, routes search results and content downloads through the wide area distribution system  
7 203 to end users, and routes billing information from the customer billing server (not shown)  
8 to the end users. The request and results processing server 300 performs the basic processing  
9 and routing related to user search requests, content download requests, administrative  
10 information requests, search results, related content suggestions, and programming  
11 notification.

12 Figure 5 shows individual subsystems of the request and results processing server  
13 300. The types of requests from a system user include, but are not limited to, programming  
14 and other content search requests, content download requests based on results of an earlier  
15 search, and administrative information requests. Requests for administrative information  
16 may be automatically generated by the user terminal 202 or may be manually initiated by the  
17 system user. When an administrative request is received, a request receiver and router 301  
18 opens a dialog with a database administrator 502 (see Figure 7) within the system  
19 administrator 500 (see Figure 8), retrieves the requested system user's administrative data,  
20 and routes the data through the communications server 250 to the user terminal 202. Search  
21 requests and content download requests are routed to a search request processor 303 and a  
22 content download request processor 302, respectively. The processors 302 and 303 open a  
23 dialog with the database administrator 502 and verify that the user's request is authorized. If  
24 the user's request is not authorized, the processor (302 or 303) sends a message to the user  
25 through the request receiver and router 301 and the communications server 250 informing the  
26 user that the request is denied. The processor (302 or 303) then logs the event with the  
27 system administrator 500. If the request is authorized, the processor (302 or 303) formats the  
28 request as necessary and routes content download requests to the content acquisition server  
29 400 and search requests to the search engine server 350.

30 In addition to searching for content containing criteria entered by a system user, a  
31 content search suggestion engine 304, in conjunction with the search engine server 350 will

be able to suggest content to the user that is related in various ways, such as by category or theme. The content search suggestion engine 304 is shown in detail in Figure 6a. For example, if a user wishes to see programs about *Titanic*, the content search suggestion engine 304 may, in addition to suggesting programs about *Titanic*, suggest or inform the user of programs and other content such as electronic books about ships other than *Titanic*. Likewise, if the search criteria include Johnny Weismuller, an actor who starred in Tarzan movies, the content search suggestion engine 304 might suggest programs and other content about Tarzan featuring someone other than Johnny Weismuller. Furthermore, the content search suggestion engine 304 may suggest programs for viewing based on past search criteria entered by the user as well as information on content the user has actually downloaded. For example, if the search criteria includes Johnny Weismuller and the user has searched and/or downloaded numerous sports-related programming in the past, the content search suggestion engine 304 may suggest programming and other content including swimming competitions and sports history and biography programming as well as Tarzan movies and other content directly related to Johnny Weismuller such as the Jungle Patrol television series. If the user searched for and received Tarzan movies, the programming search suggestion engine 304 might suggest electronic books by Edgar Rice Burroughs. Such electronic books could then be downloaded to the user terminal 202 using the wide area network/Intranet 205 bypassing the aggregator 201, or could be compiled at the aggregator 201 for delivery to the user terminal 202. Electronic book delivery systems are described in U.S. Patent 5,986,690, entitled "Electronic Book Selection and Delivery System," to Hendricks, and in co-pending patent application serial numbers 09/289,957, entitled "Electronic Book Alternative Delivery Systems," and 09/289,956, entitled "Electronic Book Alternative Delivery Methods," both of which are hereby incorporated by reference.

During submission of a content search request, the user may have the option of disabling the return of suggested content results if desired. In order to retrieve results for suggested programming, the content search suggestion engine 304 analyzes a search request form (see Figures 11a-11c) received from the user terminal 202, and performs a search within a suggestion database 308 for additional search criteria to be used to find suggested content related in some way to the original search request results. These suggested search criteria are appended to the search request form 618 forming an augmented search request

1 form (not shown) when the search request form 618 is forwarded to the search engine server  
2 350 in order for a suggested content list to be returned. U.S. Patent No. 5,798,785 entitled  
3 "Terminal for Suggesting Programs Offered on a Television Program Delivery System," and  
4 co-pending patent application Serial Number 08/906,469, filed August 5, 1997, entitled  
5 "Reprogrammable Terminal for Suggesting Programs Offered on a Television Program  
6 Delivery System," both of which are hereby incorporated by reference, provide additional  
7 details on program suggestion.

8 The suggestion database processor 307 (shown in Figure 6a and in more detail in  
9 Figure 6c) conducts a search of the suggestion database 308 to find related search criteria to  
10 use to locate suggested content. The processor 307 also performs management functions for  
11 the other subsystems of the content search suggestion engine 304. When prompted by the  
12 suggestion database processor 307, the content metadata crawler 309 crawls, or intensively  
13 searches, the metadata in the aggregator local database 501. The suggestion database  
14 processor 307 may prompt the content metadata crawler 309 on a periodic basis, for example  
15 daily. Alternatively, the prompt may be based on the occurrence of a specific event, such as  
16 an update of the aggregator local database 501. In an embodiment, the suggestion database  
17 processor 307 does not provide a prompt, and the content metadata crawler 309 continually  
18 crawls the metadata files in the local aggregator database 501.

19 The content metadata crawler 309 passes a listing of all content metadata in the  
20 aggregator local database 501 to the suggestion keyword indexer 310. The suggestion  
21 keyword indexer 310 then indexes each metadata element. (A content metadata element may  
22 be a term contained in a record that defines or describes a program or content file.) The  
23 suggestion keyword indexer 310 may assign each metadata element with a vector quantity  
24 that describes the metadata element as a suggestion keyword associated with a particular  
25 content entity. In an embodiment, the vector quantity is a measure of similarity between two  
26 words. The vector quantity may be calculated by determining relative values of several  
27 categorized descriptions of each word, including word types, parts of speech of each word,  
28 and relative locations or linkages in a thesaurus and/or lexicon. The indexer 310 then  
29 populates the suggestion database 308 with these indexed suggestion keywords, which are  
30 subsequently used by the suggestion database processor 307 along with historical data  
31 relevant to the user, to supply a list of suggested search terms. The search terms are then



added to the search request form 618 to create the augmented search request form that is returned to the search request processor 303 (see Figure 5).

The suggestion keyword indexer 310, shown in detail in Figure 6b, includes an extraction module 321 that extracts and caches a value of each metadata field. Metadata fields may be considered to fall within one or more categories such as uniquely identifying and non-uniquely identifying fields. Non-uniquely identifying fields may be those fields that classify the content, and may include genre, category, subject, and content format, for example. The uniquely identifying fields may include program content information, or metadata, including one or more of content type, content title, date of production, rating and parental notice information, performer, artist, writer, author, plot summary, keyword list, and textual content description. Coupled to the extraction module 321 is a parsing module 323 that parses contents of uniquely identifying metadata fields. The contents of a uniquely identifying field may include one or more word items that may be used to suggest additional content for access by the user. Coupled to the parsing module 323 is a classifying module 325 that may classify one or more of the word items. A comparison module 327 coupled to the classifying module 325 may compare word items to a set of database entries to determine a list of related terms. An index matrix record builder 329 may then create and augment an index matrix record for each of the classified word items.

Coupled to the comparison module 327 may be one or more databases, such as a dictionary database 331, a thesaurus database 333 and a lexicon database 335. The comparison module 327 compares a word item to entries in one or more of the dictionary database 331, the thesaurus database 333 and the lexicon database 335 and produces, as the list of related terms, one or more of a dictionary definition, lexicon data, and one or more synonyms.

The classifying module 325 may include computational linguistics tools, such as a rule-based part-of-speech tagging algorithm 337 and a stochastic part-of-speech tagging algorithm 339. The classifying module 325 uses the computational linguistic tools to determine part-of-speech data of a word item. The index matrix record builder 329 may then add the part-of-speech data to the index matrix record for the word item.

1 The index matrix record builder 329 may include a vector assignment module 341  
2 that assigns a word item vector value for a word item. The word item vector value may be a  
3 measure of similarity between a word item and a related term.

4 Figure 6c shows the suggestion database processor 307 in more detail. A vector  
5 determination module 343 assigns a search term suggestion vector range to one or more of  
6 the search request criteria contained in the search request form 618 (See Figures 11a – 11c).  
7 A vector value comparator 345 compares the search term suggestion vector range and the  
8 word item vector value to determine if a word item vector value falls within the suggestion  
9 vector range. Word items that fall within the suggestion vector range may be used to search  
10 the suggestion database.

11 A database administrator 502, shown in detail in Figure 7, as part of the system  
12 administrator 500 (see Figure 8) maintains the aggregator local database 501 and interfaces  
13 with the various components of the aggregator 201 that retrieve information from and store  
14 information to the aggregator local database 501. Word items that fall within the suggestion  
15 vector range may be appended to a search request form (see Figures 11a – 11b) to create an  
16 augmented search request form that may be used by the search request processor 303 to  
17 return suggested content results. A group of individual databases maintained within the  
18 database administrator 502 constitutes the aggregator local database 501. An administrative  
19 database server 509 maintains an aggregator administrative database 510, which stores and  
20 processes information including, but not limited to, authorized system administrators,  
21 passwords, and administrator usage rights. The administrative database server 509 also  
22 governs replication of all databases to aggregator local storage 254 and backup of aggregator  
23 local storage 254 to the aggregator archives 255. A user database server 511 maintains an  
24 aggregator user database 512, which stores and processes information including, but not  
25 limited to, user account data, user profile information, user subscription services, user access  
26 rights, and past user search and download data (if authorized by the user).

27 A content provider database server 513 maintains a content provider database 514,  
28 which stores and processes information including, but not limited to, a roster of registered  
29 content providers and their administrative data, remote provider administrators and their  
30 passwords and access rights, and terms of agreements between providers and the operator of  
31 the particular aggregator 201. A content database server 515 maintains an aggregator local

1 content database 516, an aggregator remote content database 517, and an aggregator  
2 scheduled content database 518. The aggregator local content database 516 stores and  
3 processes information including, but not limited to, the catalog of all content stored in the  
4 aggregator local storage 254 and all metadata associated with that content. The aggregator  
5 remote content database 517 stores and processes information including, but not limited to,  
6 the catalog of all content stored in each remote database 261 and all metadata associated with  
7 that content. The aggregator scheduled content database 518 stores and processes  
8 information including, but not limited to, scheduled broadcast programming content and  
9 scheduled streaming program content available to users, available live programming,  
10 recurring scheduled programming, and program schedule metadata information.

11 The system user can initiate content search requests by entering several different  
12 types of search criteria using several different options of input device or method. Search  
13 criteria can be entered using keywords that relate to certain aspects of programming content  
14 including, but not limited to, subject, author, title, cast members or performers, director,  
15 and/or content description. A search can be based on program type or format including, but  
16 not limited to, movies, television series, documentaries, sports programs, talk radio, music  
17 radio, and electronic books. The user may input search criteria based on subject categories  
18 including, but not limited to, action, drama, history, educational, juvenile, adult, current  
19 events, nature, live events, and music categories such as classical, jazz, rock, consumer  
20 affairs, political content, and geographically specific content. Search criteria can also be  
21 entered based on time of day, channel, and/or content provider. The user may also input free  
22 form questions using plain speech patterns. In any case, the different search criteria can be  
23 used separately or in conjunction with one another to refine a content search. An example  
24 might be the entry of “*Titanic*” as a keyword, “Documentary” for program type, and the free  
25 form question “What company built *Titanic*?”.

26 The content search suggestion engine 304 shown in Figures 6a-6c builds and  
27 maintains the suggestion database 308 and analyzes content search requests along with the  
28 users’ histories and profiles to create lists of suggested search terms to be appended to the  
29 search request form 618. Content is suggested by creating a list of suggested search criteria  
30 based on the user’s profile, history, and current search request and submitting this suggested  
31 search criteria in parallel with the user defined search criteria to the search engine server 350.

1 The content metadata crawler 309 performs a crawl, or thorough search, of the content  
2 metadata contained in the aggregator local database 501 maintained by the system  
3 administrator 500. In an embodiment, the search suggestion engine 304 may contain and  
4 maintain a replica of the content portion of the aggregator local database 501 in order to  
5 perform metadata crawling internally thereby reducing the processor and communications  
6 loading on the system administrator 500. Figure 9 is an overview of the subroutines that  
7 build and maintain the suggestion database 308. The process is initiated with routine 101,  
8 which receives instructions from the suggestion database processor 307 to begin a metadata  
9 crawling operation on the aggregator local database (see Figure 8). Routine 101 may then  
10 open a communications session with the database administrator 502. Routine 102 then  
11 queries the next record in the database and routine 103 tests whether an end-of-file condition  
12 within the database has been reached. If the end of file condition has been met, the metadata  
13 crawling process has accessed every record in the local aggregator database and processing is  
14 passed to routine 113 which reports administrative data back to the suggestion database  
15 processor 307 and terminates the metadata crawling process, routine 114. Until the end of  
16 file condition has been met, processing is passed from routine 103 to routine 104, which  
17 reads the entire metadata entry for the current content record into local memory.

18 Processing is then routed to the suggestion keyword indexer 310 module, which  
19 extracts and caches the values, or entries, of each metadata field, routine 105. Routine 106  
20 parses the contents of uniquely identifying metadata fields by breaking down the entries into  
21 component words and phrases and eliminating common terms and words such as “the,”  
22 “and,” “it,” “which,” “then,” “by,” “of,” “he,” “she,” and “them.” Uniquely identifying  
23 metadata fields in the metadata record may be those metadata fields that do not have a finite  
24 list of potential values. Examples of uniquely identifying fields include content title,  
25 performers, artist, writer, author, keyword list, plot summary, and textual content description.  
26 Routine 106 then caches, or saves, each word item into temporary local memory, or cache,  
27 creating a temporary list of word items. Routine 107 then reads the next cached word item  
28 and routine 108 tests whether the end of the cached list of terms has been reached, in which  
29 case the processing is routed back to routine 102 to access the next metadata record in the  
30 database. Until the end of the cached list is reached, processing is passed to routine 109,  
31 which determines if an index matrix record already exists for the current word item. If an

1 index matrix record does not already exist, routine 110 creates a new index matrix record for  
2 the word item in the suggestion database 308 and then routes processing to routine 111.  
3 Routine 111 determines the part of speech of the word item, such as noun or verb, using  
4 computational linguistics tools that may include ruled-based and / or stochastic parts-of-  
5 speech tagging algorithms. Routine 111 may also access one or more thesaurus and / or  
6 lexicon databases and return data that may include the word item's definition and a list of  
7 related terms including synonyms. One or more thesaurus and/or lexicon databases may  
8 reside locally in the suggestion database 308 or elsewhere within the system 200 or may  
9 reside remotely on the network and may be accessed through the network gateway 251.  
10 Routine 111 then adds part-of-speech, lexicon, and thesaurus data to the word item's index  
11 matrix record. Words that cannot be tagged as known parts of speech may be designated as  
12 possibly being proper nouns or names and are verified against known proper nouns and  
13 names located in the lexicon databases and this data is added to the word item's index matrix  
14 record. The suggestion keyword indexer 310 may add unknown words to the index matrix  
15 within the suggestion database 308 and may track occurrences of unknown word items in  
16 order to establish recurring word items as desirable entries in the index matrix and to  
17 gradually downgrade and eliminate word items that in fact do not have relevance to digital  
18 content being made available by the system 200.

19 Processing then continues within the suggestion keyword indexer 310 with routine  
20 112. If it was determined earlier by routine 109 that an index matrix record already existed  
21 for the current word item, then processing would have been routed directly to routine 112  
22 bypassing routines 110 and 111, which would have already been completed for the current  
23 word item. Metadata that was cached by routines 105 and 106 for the current word item is  
24 now entered into the index matrix record for the current word item, routine 112. The index  
25 matrix comprises numerous tabular entries or records similar to the structure of a database  
26 and each entry may contain numerous fields with field descriptive information and field entry  
27 data that may assume various alphanumeric values. The values that a field entry may assume  
28 may be either unlimited or may be limited to a finite list of potential values. Algorithms may  
29 operate on index matrix entries as vector quantities either in whole or in part. In an  
30 embodiment, the suggestion database 308 comprises the metadata index matrix that contains  
31 a matrix entry for every word item parsed from all available metadata records in the

1 aggregator local database. The metadata index matrix may be structured so that every word  
2 item has an index matrix record that contains information about that word item such as the  
3 original metadata field where the item was found and a listing of all other word items found  
4 in all other fields. Through the iterative metadata crawling and indexing process 100 shown  
5 in Figure 9, the suggestion database 308 may be populated with virtually every word, term,  
6 and name associated with all types of digital content to be made available to system 200  
7 users.

8 Figure 10 shows an overview of the suggested search criteria retrieval process 120  
9 performed by the suggestion database processor 307 in which the current search request  
10 criteria along with user profile and history data may be used as comparison criteria for  
11 retrieving suggested search parameters from the suggestion database 308. The content search  
12 suggestion engine 304 first receives the search request criteria (routine 121) for the current  
13 user search request from the search request processor 303 within the request and results  
14 processing server 300 and puts the search requests criteria into a vector format like that of the  
15 index matrix records. The suggestion engine 304 then opens a communications dialog with  
16 the database administrator 502 within the system administrator 500 and requests and retrieves  
17 the user profile and user history analysis report for the user that has submitted the current  
18 search request, routine 122. The user profile comprises content search parameters and  
19 preferences that may be controlled by the user and may be augmented by the system  
20 administrator 500 based on criteria such as user account type, user access level, and historical  
21 data. The user history analysis report is a summary of statistical analysis of the user's  
22 previous account activities that may include previous search requests, returned search results,  
23 and content download requests. The history analysis report is compiled by the user database  
24 server 511 within the database administrator 502 and comprises information such as a  
25 percentage breakdown of content types and genres for which a user has requested searches,  
26 statistical breakdown of content that the user has downloaded, and summary data about  
27 content that was selected for download from earlier suggested content results.

28 The suggestion database processor 307 then takes the vector-formatted search request  
29 criteria and performs a comparison against the index matrix, routine 123. Each comparison  
30 may determine whether a particular index matrix record falls within a suggestion vector  
31 range of the search criteria, and if so routine 124 caches that particular matrix record into a

temporary memory list. The suggestion vector range may be a measure of how similar the matrix record is to the search criteria and may have its values be adjustable so that tighter ranges may result in fewer, more-relevant suggestion criteria and wider ranges may result in a larger number of somewhat broader results. This suggestion vector range may be adjustable by the user in the form of scaling how wide or how narrow the user wishes suggested search results to be and may also be adjustable automatically by the suggestion database processor 307 based on trend analysis of the user's previous download of suggested content. The vector quantity of a search request criteria and an index matrix record is a measure of similarity between the two entries that is calculated by determining relative values of several categorical descriptions of each word including word type, part of speech of each word, relative locations or linkages in a thesaurus, and occurrence of a search term within a relevant or similar term field of an index matrix record.

Once the comparison has been completed for the entire index matrix, the cached records that fell within the suggestion vector range are then filtered by routine 125 according to the user's profile. This filtering may take into account such factors as the user's preferences and approved content access level. The filtered results are then forwarded to routine 126, which then ranks the results according to the user's historical analysis report and ranks the results by relevance and by similarities to previously searched and downloaded content. The ranking of results may take into account historical criteria so that if a user consistently searches more frequently by selecting program type, then matching program type suggestions will be weighted more and therefore ranked higher in the suggested search criteria than other attributes such as genre or content provider. The resulting list of suggested search criteria is then formatted into a data file and sent by routine 127 to the search request processor 303 for inclusion in the content search. Processing then stops at routine 128.

The content suggestion process may be carried out without intervention by or the attention of the system user so that the process may be transparent to the user with the exception that content results may contain a section designated as suggested content or individual content listings may be designated as suggested versus being a direct result of search criteria originally input by the user. This system user may be given the option to control the level of suggested content that is returned with search results and may range from no suggested results up to a broad list of suggested content results. If desired, the user may

- 1 also be able select or filter user profile and historical data that may be used for the search
- 2 criteria suggestion process.

Prepared